Testing for the Cointegrating Rank of a VAR Process With Structural Shifts

Pentti SAIKKONEN
Department of Statistics, University of Helsinki, FIN-00014 University of Helsinki, Finland (*saikkone@valt.helsinki.fi*)

Helmut LÜTKEPOHL
Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-University, 10178 Berlin, Germany (*luetke@wiwi.hu-berlin.de*)

Tests for the cointegrating rank of a vector autoregressive process are considered that allow for possible exogenous shifts in the mean of the data-generation process. The break points are assumed to be known a priori. It is proposed to estimate and remove the deterministic terms such as mean, linear-trend term, and a shift in a first step. Then systems cointegration tests are applied to the adjusted series. The resulting tests are shown to have known limiting null distributions that are free of nuisance parameters and do not depend on the break point. The tests are applied for analyzing the number of cointegrating relations in two German money-demand systems.

KEY WORDS: Cointegration; Money-demand analysis; Testing for cointegration; Vector autoregressive process.

Many economic time series exhibit breaks or shifts in their levels that are not consistent with standard types of data-generation processes (DGP's). Such breaks are often caused by exogenous events that have occurred during the observation period. For example, the German unification has caused shifts in several macroeconomic time series such as gross national product (GNP) and measures of the money stock. In this example the timing and the reasons for the shifts are known. In other situations, neither the timing nor whether a shift actually has occurred are known at the outset of an analysis.

Corresponding to the various types of structural changes and problems related to them, there is an extensive literature dealing with the consequences of structural shifts for estimation and testing procedures in univariate and multivariate time series models as well as in regression models for time series variables [e.g., see Hackl and Westlund (1989) for a large number of references to the earlier literature]. In particular, in many studies testing for unit roots and breaks in univariate time series is considered. Examples are Perron (1989, 1990), Perron and Vogelsang (1992), Rappoport and Reichlin (1989), Zivot and Andrews (1992), Banerjee, Lumsdaine, and Stock (1992), Amsler and Lee (1995), and Ghysels and Perron (1996) to name just a few. In these works, different assumptions regarding the DGP are made. For instance, the break point may be known or unknown, it may be a shift in the level of a series, or it may be a break in the deterministic trend component. Structural shifts in the context of cointegration analysis were considered, for example, by Quintos (1998), Seo (1998), Hansen (1992), Gregory and Hansen (1996), Campos, Ericsson, and Hendry (1996), Johansen and Nielsen (1993), and Inoue (1999), among others. Quintos and Seo focused on changes in the cointegration or adjustment parameters, whereas Hansen, Gregory and Hansen, and Campos et al. discussed tests for cointegration in a single-equation framework. In contrast, Johansen and Nielsen (1993) studied the consequences of structural breaks in a systems context and derived likelihood ratio (LR) tests for the number of cointegrating relations in a system of variables. Inoue (1999) also considered testing for the cointegrating rank of a system. He assumed, however, that the break is present under the alternative only.

The overall message from these studies is that structural breaks can distort standard inference procedures substantially and, hence, it is necessary to make appropriate adjustments if structural shifts are known to have occurred or are suspected. If there is just one break point in the observation period, it may be tempting to analyze the two regimes before and after the shift separately. This may, however, result in a substantial loss in efficiency and/or power. Therefore, procedures that take structural changes into account by adjusting the inference methods are often preferable to eliminating parts of the sample.

In this study we will consider the problem of how to test for the number of cointegrating relations in a system of variables if some of them have a shift in the mean at a known time point. This situation is relevant, for instance, in the aforementioned case of German data if the sampling period covers the German unification. Assuming that the DGP is a finite-order vector autoregressive (VAR) process with a shift in the mean, tests for the number of cointegrating relations will be proposed with an asymptotic null distribution that is free of nuisance parameters and does not depend on where the break point has occurred. In contrast to the assumption of Inoue (1999), the shift is allowed to be present both under the null and under the alternative hypothesis. The fact that the null distribution does not depend on the break point contrasts with the tests analyzed by Johansen and Nielsen (1993), whose asymptotic null distribution de-

Journal of Business & Economic Statistics
October 2000, Vol. 18, No. 4

pends on when the shift has occurred. If the break point is assumed to be known, this does not necessarily mean that the tests depend on unknown nuisance parameters. It means, however, that new critical values have to be obtained for each specific situation. Even if just a few new observations become available, generating new critical values will generally be necessary to perform the tests. In contrast, the tests proposed in the following have a known asymptotic distribution that does not depend on the break date so that no new simulations are required. Tables with critical values are available elsewhere in the literature. Moreover, our tests can be adopted for time series with more than one shift in the mean or to series with single outlying observations.

The idea underlying our new tests is to estimate and remove the deterministic parts including the shifts in a first step and then perform a test for the cointegrating rank on the adjusted series. The deterministic part of the DGP may, in fact, include a linear time trend in addition to shifts in the mean term. Similar ideas were used by Amsler and Lee (1995) in unit-root tests in the presence of structural breaks and by Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (in press) to construct tests for VAR processes with deterministic linear trends and without structural shifts. Because we will refer to the latter two articles several times in the following, we will abbreviate them as L&S and S&L, respectively.

The structure of the article is as follows. In Section 1, the DGP is precisely specified and the assumptions underlying our analysis are laid out. The estimation of the parameters of the deterministic parts is discussed in Section 2, and the cointegration tests are presented in Section 3. A simulation study exploring the small-sample properties of our new tests is described in Section 4, and examples based on German macro data are given in Section 5. Conclusions follow in Section 6. The proofs of our theorems are provided in Appendix A, and the sources of the data used in the empirical example are given in Appendix B.

The following general notation is used. The lag and differencing operators are denoted by L and Δ, respectively; that is, for a time series or stochastic process y_t we have $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $\mathrm{I}(d)$ denotes an integrated process of order d; that is, the process is stationary or asymptotically stationary after differencing d times but it is still nonstationary after differencing just $d-1$ times. Convergence in distribution or weak convergence is signified by $\xrightarrow{d}$. The trace and the rank of the matrix A are denoted by $\mathrm{tr}(A)$ and $\mathrm{rk}(A)$, respectively. Moreover, $\|\cdot\|$ denotes the Euclidean norm. If A is an $(n \times m)$ matrix of full column rank $(n > m)$, we denote an orthogonal complement by $A_\perp$ so that $A_\perp$ is an $(n \times (n-m))$ matrix of full column rank and such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is 0, and the orthogonal complement of a zero matrix is an identity matrix of suitable dimension. An $(n \times n)$ identity matrix is denoted by I_n. LS, GLS, and RR are used to abbreviate least squares, generalized least squares, and reduced rank, respectively. LR and LM tests are short for likelihood ratio and Lagrange multiplier tests. DGP stands for data-generation process, and ECM abbreviates error-correction model. A sum is defined to be 0 if the lower bound of the summation index exceeds the upper bound.

1. THE MODEL

Suppose an observed n-dimensional time series $y_t = (y_{1t}, \ldots, y_{nt})'$ $(t = 1, \ldots, T)$ is generated by the following mechanism:

$$y_t = \mu_0 + \mu_1 t + \delta_0 d_{0t} + \delta_1 d_{1t} + x_t, \qquad t = 1, 2, \ldots, \tag{1.1}$$

where μ_i and δ_i $(i = 0, 1)$ are unknown $(n \times 1)$ parameter vectors and d_{0t} and d_{1t} are dummy variables defined as

$$d_{0t} = \begin{cases} 1, & t = T_0 \\ 0, & t \neq T_0 \end{cases} \tag{1.2}$$

and

$$d_{1t} = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases}; \tag{1.3}$$

that is, d_{0t} is an impulse dummy and d_{1t} is a step dummy variable. The associated terms allow taking into account sudden changes in the mean of the process that occur, for instance, in German macroeconomic time series at the time of the reunification.

The term x_t is an unobservable error process that is assumed to have a VAR(p) representation,

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t, \quad t = 1, 2, \ldots. \tag{1.4}$$

Here the A_j are $(n \times n)$ coefficient matrices and, for simplicity, we assume that $x_t = 0$ for $t \leq 0$ and $\varepsilon_t \sim \mathrm{NID}(0, \Omega)$; that is, the ε_t are independent, identically distributed (iid) Gaussian vectors with zero mean and covariance matrix Ω. The normality assumption is made here for convenience although it could be replaced by an iid assumption and sufficient moment conditions. The assumption regarding the initial values is also a convenient simplification. The asymptotic results derived later remain valid if the initial values are assumed to be from a fixed probability distribution that does not depend on the sample size. The usual error-correction form of (1.4), obtained by subtracting x_{t-1} from both sides of (1.4) and rearranging terms, is given by

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1.5}$$

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ $(j = 1, \ldots, p-1)$ are $(n \times n)$ matrices.

We assume that x_t is at most I(1) and cointegrated with cointegrating rank r. This implies in particular that the Granger representation theorem of Johansen (1995, chap. 4) is assumed to hold. Hence, the matrix Π can be written as

$$\Pi = \alpha\beta', \tag{1.6}$$

where α and β are $(n \times r)$ matrices of full-column rank. As is well known, $\beta' x_t$ and Δx_t are then zero mean (asymptotically) stationary processes. Moreover, defining

$$\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} j A_{j+1}$$

and $C = \beta_\perp (\alpha'_\perp \Psi \beta_\perp)^{-1} \alpha'_\perp$, we have

$$x_t = C \sum_{j=1}^{t} \varepsilon_j + \xi_t, \qquad t = 1, 2, \ldots, \tag{1.7}$$

where ξ_t is a zero mean (asymptotically) stationary process.

Now we consider the dummy variables d_{0t} and d_{1t}. We assume that the values of the integers T_0 and T_1 are known a priori and, if y_t is observed for $t = 1, \ldots, T$, then $T_0 \leq T$ and $T_1 \leq T$. The case $T_0 = T_1$ is possible unless $T_0 = T_1 = T$. We exclude this latter possibility by assuming $T_1 < T - p$. It is also convenient to assume that $T_0 > p$ and $T_1 > p$ and, furthermore, that

$$\lim_{T \to \infty} \frac{T_1}{T} = a_1 \quad \text{with} \quad 0 < a_1 \leq 1. \tag{1.8}$$

In other words, the break point T_1 may be thought of as occurring at a fixed proportion of the full sample size even if an asymptotic analysis is performed where $T \to \infty$. Alternatively, the break may be viewed as having occurred a fixed number of periods before the end of the sample period. In the latter case, $a_1 = 1$. These assumptions for the integers T_0 and T_1 are not restrictive from a practical point of view. This is obvious for the aforementioned inequalities. Condition (1.8) may look somewhat restrictive because it implies that the jump in the dummy d_{1t} is not allowed to take place at the beginning of the sample so that, for example, T_1 would be only slightly larger than p. This would mean that the parameters μ_0 and δ_0 would become asymptotically indistinguishable. This problem can readily be handled by redefining d_{t1}, however, so that it takes the value 1 for $t < T_1$ and 0 for $t \geq T_1$. Then the inequalities in (1.8) should be changed to $0 \leq a_1 < 1$, and it is easy to see that our subsequent derivations apply with only minor and fairly obvious modifications. In summary, the preceding discussion shows that the jump in the dummy variable d_{1t} can also take place at the beginning or at the end of the sample but, for ease of exposition, we exclude one of these two possibilities. Thus, our assumptions about the integers T_0 and T_1 are quite general and weaker than in some previous studies in which condition (1.8) is required with $0 < a_1 < 1$. Finally, note that we assume for convenience and for expository purposes that there is just one impulse dummy d_{0t} and one step dummy d_{1t}. It is not difficult to see that our results can be adapted to models with any number of (linearly independent) impulse dummies and step dummies. Moreover, our results can easily be modified to the case in which the model contains only step dummies or impulse dummies. It is also possible to exclude the trend term from the model; that is, $\mu_1 = 0$ may be assumed a priori. This case will be briefly discussed later on.

For our analysis, it is convenient to define the lag polynomial

$$\begin{aligned} A(L) &= I_n - A_1 L - \cdots - A_p L^p \\ &= I_n \Delta - \Pi L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1} \Delta L^{p-1} \end{aligned} \tag{1.9}$$

and notice that the relation between the two representations is given by

$$\begin{aligned} A_1 &= I_n + \alpha\beta' + \Gamma_1 \\ A_j &= \Gamma_j - \Gamma_{j-1}, \qquad j = 2, \ldots, p-1 \\ A_p &= -\Gamma_{p-1}, \end{aligned} \tag{1.10}$$

where Π is expressed as in (1.6). Multiplying (1.1) by $A(L)$ yields

$$\begin{aligned} \Delta y_t &= \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \delta_0 d_{0t} \\ &\quad - \sum_{j=1}^{p} A_j \delta_0 d_{0,t-j} + \delta_1 \Delta d_{1t} - \sum_{j=1}^{p-1} \Gamma_j \delta_1 \Delta d_{1,t-j} \\ &\quad - \Pi \delta_1 d_{1,t-1} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \end{aligned}$$

where $\nu_0 = -\Pi\mu_0 + (\Psi + \Pi)\mu_1$ and $\nu_1 = -\Pi\mu_1$. To be able to write this model in a simpler form, define

$$\gamma_{0j} = \begin{cases} \delta_0, & j = 0 \\ -A_j \delta_0, & j = 1, \ldots, p \end{cases}$$

and

$$\gamma_{1j} = \begin{cases} \delta_1, & j = 0 \\ -\Gamma_j \delta_1, & j = 1, \ldots, p-1 \end{cases}.$$

Using (1.6), we can then write

$$\begin{aligned} \Delta y_t &= \nu + \alpha(\beta' y_{t-1} - \tau(t-1) - \theta d_{1,t-1}) \\ &\quad + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p} \gamma_{0j} d_{0,t-j} + \sum_{j=0}^{p-1} \gamma_{1j} \Delta d_{1,t-j} + \varepsilon_t, \\ &\qquad t = p+1, p+2, \ldots, \end{aligned} \tag{1.11}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1, \tau = \beta'\mu_1$, and $\theta = \beta'\delta_1$. Notice that here $\Delta d_{1,t-j}$ is an impulse dummy, which takes the value 1 at $t = T_1 + j$ and 0 elsewhere.

Equation (1.11) specifies an ECM for the observed series y_t. We shall use this form of the model to obtain first-stage estimators for the parameters of the error process x_t—that is, for α, β, Γ_j $(j = 1, \ldots, p-1)$, and Ω. Some remarks on the ECM (1.11) and the estimation of its parameters are therefore in order. Using Equation (1.10) and the definitions, it can first be seen that a conventional RR regression cannot be used to obtain the ML estimators because there are nonlinear restrictions between the parameters in (1.11). To obtain the previously mentioned first-stage estimators, we shall simply ignore these restrictions. This should not cause any great loss of efficiency because the restrictions occur in coefficient vectors of impulse dummies only. Before computing the estimators, one should check, however, that the dummy variables on the right side of (1.11) are linearly independent. Because both d_{0t} and Δd_{1t} are impulse dummies, it is possible that some impulse dummies appear twice in (1.11) and can be combined. This, of course, has no effect on the estimation of the parameters α, β, Γ_j, and Ω that are of interest at this point. For simplicity, we assume that all dummy variables in (1.11) are linearly independent. The assumption $p < T_1 < T - p$ guarantees that there cannot be linear dependencies between the step dummy $d_{1,t-1}$ and the impulse dummies. Such linear dependencies could be eliminated by excluding impulse dummies so that the assumption $T_1 < T - p$ is actually not needed here. It is convenient in proving Theorem 2.1, however, and is therefore imposed because assuming that $T_1 < T - p$ seems harmless from a practical point of view.

In the framework of our model, we are interested in testing whether the assumption made for the rank of the matrix Π is correct. In other words, for some prespecified rank r_0, we wish to consider testing the null hypothesis

$$H_0(r_0): \mathrm{rk}(\Pi) = r_0 \quad \text{vs.} \quad H_1(r_0): \mathrm{rk}(\Pi) > r_0. \quad (1.12)$$

In this context it turns out to be very useful to make explicit use of the assumption that the DGP is of the form (1.1). As will be seen later, it is then possible to obtain tests with convenient asymptotic properties. As mentioned in the introduction, our formulation of the model allows us to estimate the deterministic part of the DGP first and then apply cointegration tests to the process adjusted for deterministic terms. In Section 2, estimators of the parameters of the deterministic part will be presented, and the cointegration tests will be considered in Section 3.

2. ESTIMATING THE PARAMETERS OF THE DETERMINISTIC PART OF THE MODEL

We shall estimate the parameters μ_i and δ_i $(i = 0, 1)$ in (1.1) by using a feasible GLS approach similar to that of S&L. Note that in our model GLS is not equivalent to LS estimation of the parameters of the deterministic part despite the fact that all equations involve the same regressors. The reason is that the error term x_t is autocorrelated and may in fact consist of integrated components. This means that LS and GLS are not even asymptotically equivalent (e.g., see Xiao and Phillips 1999 and the references therein).

To derive our estimator of the parameters of the deterministic part we define

$$a_{0t} = \begin{cases} 1 & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases} \quad \text{and} \quad a_{1t} = \begin{cases} t & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases}.$$

Multiplying (1.1) from the left by $A(L)$ gives

$$A(L)y_t = H_{0t}\mu_0 + H_{1t}\mu_1 + K_{0t}\delta_0 + K_{1t}\delta_1 + \varepsilon_t, \quad t = 1, 2, \ldots, \quad (2.1)$$

where the y_t are set to 0 for $t \leq 0, H_{it} = A(L)a_{it}$, and $K_{it} = A(L)d_{it}, i = 0, 1$. As in S&L, we also define the matrix

$$Q = [\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1/2} : \alpha_\perp(\alpha'_\perp\Omega\alpha_\perp)^{-1/2}] \quad (2.2)$$

with the property $QQ' = \Omega^{-1}$. Premultiplying (2.1) by Q' transforms the covariance matrix of the error term to an identity matrix so that, as required in GLS estimation, we have a transformation that results in a (multivariate) regression model with standard properties of the error term.

To make the preceding transformation feasible, suitable estimators of the parameters α, β, Γ_j $(j = 1, \ldots, p-1)$, and Ω are needed. Such estimators can be obtained by an RR regression of (1.11) in the way discussed in Section 1. These estimators are denoted by $\tilde{\alpha}, \tilde{\beta}, \tilde{\Gamma}_j$, and $\tilde{\Omega}$. Substituting them for the corresponding theoretical parameters in (1.10) gives estimators for the coefficient matrices A_j. Denoting these estimators by $\tilde{A}_j$, we can define $\tilde{A}(L) = I_n - \tilde{A}_1 L - \cdots - \tilde{A}_p L^p$ and, furthermore, $\tilde{H}_{it} = \tilde{A}(L)a_{it}$ and $\tilde{K}_{it} = \tilde{A}(L)d_{it}$ $(i = 0, 1)$. Note that $\tilde{A}(L)$ satisfies the cointegrating restrictions. Thus, we are able to construct a feasible analog of (2.1). Moreover, a suitable estimator of the matrix Q can be readily obtained by forming $\tilde{\alpha}_\perp$ from $\tilde{\alpha}$ and replacing Ω, α, and $\alpha_\perp$ in the definition of Q by their estimators. If $\tilde{Q}$ is used to denote the resulting estimator of Q, we can finally introduce the multivariate auxiliary regression model

$$\tilde{Q}'\tilde{A}(L)y_t = \tilde{Q}'\tilde{H}_{0t}\mu_0 + \tilde{Q}'\tilde{H}_{1t}\mu_1 + \tilde{Q}'\tilde{K}_{0t}\delta_0 + \tilde{Q}'\tilde{K}_{1t}\delta_1 + \eta_t, \quad t = 1, \ldots, T. \quad (2.3)$$

The LS estimators of the parameters μ_i and δ_i will be denoted by $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$, respectively. They are used in Section 3 to obtain tests for the cointegrating rank. It is, of course, apparent that the estimator $\hat{\delta}_0$ that estimates the coefficient vector of the impulse dummy d_{0t} cannot be consistent. From S&L it is also clear that the estimator $\hat{\mu}_0$ is generally not consistent although it is consistent in the direction of β; that is, $\beta'\hat{\mu}_0$ is a consistent estimator of $\beta'\mu_0$. On the basis of this result, it is to be expected that the consistency properties of the estimators $\hat{\delta}_1$ are similar to those of $\hat{\mu}_0$ provided $T - T_1 \to \infty$. The asymptotic properties of the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ are given in the following theorem, whose proof requires suitable consistency results of the estimators $\tilde{\alpha}, \tilde{\beta}, \tilde{\Gamma}_j$, and $\tilde{\Omega}$. These results are first stated in the following lemma.

Lemma 2.1. Suppose that the assumptions made in Section 1 hold and the null hypothesis $H_0(r_0)$ is true. Suppose further that, if (1.8) holds with $a_1 = 1$, then $T - T_1$ either tends to infinity or converges to a finite constant. Define the (infeasible) estimators $\tilde{\beta}_\xi = \tilde{\beta}(\xi'\tilde{\beta})^{-1}$ and $\tilde{\alpha}_\xi = \tilde{\alpha}\tilde{\beta}'\xi$, where $\xi' = (\beta'\beta)^{-1}\beta'$. Then $\tilde{\beta}_\xi = \beta + O_p(T^{-1}), \tilde{\alpha}_\xi = \alpha + O_p(T^{-1/2}), \tilde{\Gamma}_j = \Gamma_j + O_p(T^{-1/2})$ $(j = 1, \ldots, p-1)$, and $\tilde{\Omega} = \Omega + O_p(T^{-1/2})$.

The lemma is proven in the Appendix. The results are similar to those one obtains from a model without dummy variables. They can be used to show that the same consistency properties apply with any normalization of the estimators (cf. Johansen 1995, p. 184). In the present context, the normalization does not matter, however, because we use the estimators $\tilde{\alpha}$ and $\tilde{\beta}$ in situations that are invariant to a particular normalization. Now we can state the properties of the estimators of the parameters of the deterministic part of (1.1). Again the proof is sketched in the Appendix.

Theorem 2.1. Under the conditions of Lemma 2.1,

$$\beta'(\hat{\mu}_0 - \mu_0) = O_p(T^{-1/2}), \tag{2.4}$$

$$\beta'_\perp(\hat{\mu}_0 - \mu_0) = O_p(1), \tag{2.5}$$

$$\hat{\delta}_0 - \delta_0 = O_p(1), \tag{2.6}$$

$$\beta'(\hat{\delta}_1 - \delta_1) = O_p((T - T_1)^{-1/2}), \tag{2.7}$$

$$\beta'_\perp(\hat{\delta}_1 - \delta_1) = O_p(1), \tag{2.8}$$

$$\beta'(\hat{\mu}_1 - \mu_1) = O_p(T^{-3/2}), \tag{2.9}$$

and

$$T^{1/2}\beta'_\perp(\hat{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta'_\perp C\Omega C'\beta_\perp). \tag{2.10}$$

Here $C = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp$ as before and all quantities converge jointly in distribution on appropriate standardization.

The properties of the estimators $\hat{\mu}_0$ and $\hat{\mu}_1$ are entirely similar to those obtained by S&L, theorem 1. Whereas $\hat{\mu}_1$ is consistent, the same is not true for $\hat{\mu}_0$. The latter is consistent only in the direction of β and not in the direction of $\beta_\perp$. Even in the direction of $\beta_\perp$, however, the estimator $\hat{\mu}_0$ is bounded in probability and this property turns out to be sufficient for our purposes. As to the estimators $\hat{\delta}_0$ and $\hat{\delta}_1$, the obvious inconsistency of the former is implied by (2.6), while (2.7) and (2.8) show that the asymptotic behavior of the latter is similar to that of $\hat{\mu}_0$ when $T - T_1 \to \infty$ or, in other words, when information on the parameter δ_1 in the direction of β increases with the sample size. A general conclusion from (2.7) and (2.8) is that the estimator $\hat{\delta}_1$ is never consistent in the direction of $\beta_\perp$ and it can also be inconsistent in the direction of β if $T - T_1$ does not go to infinity. As in the case of $\hat{\mu}_0$, from our point of view it is important that the estimator $\hat{\delta}_1$ is bounded in probability, particularly in the direction of $\beta_\perp$. We will now discuss how these estimators can be used in constructing tests for the pair of hypotheses in (1.12).

3. TEST PROCEDURES

When the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$ are available, one can form a sample analog of the series x_t as $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t - \hat{\delta}_0 d_{0t} - \hat{\delta}_1 d_{1t}$ and use it to obtain LM-type or LR-type test statistics for the hypothesis $H_0(r_0)$ in the same way as L&S and S&L. The LM-type test statistic requires estimators of the parameters α, β, and Ω. The RR regression estimators based on (1.11) and discussed in Section 2 can be used for this purpose. Alternatively, the LR-type test statistic may be obtained in the same way as the usual LR test statistic from the feasible counterpart of the ECM (1.5); that is, it is determined from

$$\Delta\hat{x}_t = \Pi\hat{x}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j \Delta\hat{x}_{t-j} + e_t, \qquad t = p+1, \ldots, T, \tag{3.1}$$

where e_t is an error term defined explicitly in the Appendix. The following general formulation discusses LM- and LR-type test statistics obtained in both of the aforementioned ways.

Let $\tilde{\alpha}, \tilde{\beta}$, and $\tilde{\Omega}$ be any estimators of the parameters α, β, and Ω, respectively, satisfying the consistency results in Lemma 1.1. Consider the auxiliary regression model

$$\tilde{\alpha}'_\perp \Delta\hat{x}_t = \varphi^* \hat{u}_{t-1} + \rho^* \hat{v}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j^* \Delta\hat{x}_{t-j} + \tilde{\alpha}'_\perp e_t^*, \qquad t = p+1, \ldots, T, \tag{3.2}$$

where $\hat{u}_t = \tilde{\beta}'\hat{x}_t, \hat{v}_t = \tilde{\beta}'_\perp \hat{x}_t$, and $e_t^* = e_t - \alpha(\tilde{\beta} - \beta)'\hat{x}_{t-1}$. The LM-type test statistic of S&L is obtained by forming the usual LM test statistic of the multivariate linear model for the null hypothesis $\rho^* = 0$ in (3.2). If the usual LR test statistic of the multivariate linear model is used instead, an asymptotically equivalent test statistic is obtained. In general this test statistic is not the LR test statistic based on (3.1). As shown by Saikkonen and Lütkepohl (1999), however, this is actually the case if $\tilde{\alpha}, \tilde{\beta}$, and $\tilde{\Omega}$ are RR regression estimators based on (3.1). Thus, to be able to use the preceding general framework to construct the LR-type test based on (3.1), we have to show that the consistency results of Lemma 2.1 hold for the RR regression estimators of α, β, and Ω obtained from (3.1). This will be done in the Appendix. Thus, testing the null hypothesis $\rho^* = 0$ in (3.2) by conventional methods of the multivariate linear model gives both the LM-type and the LR-type tests discussed by S&L. For convenience, the following test statistic assumes an LM-type (or Wald-type) form, but the LR-type form is, of course, asymptotically equivalent under the null hypothesis.

Following S&L, we now introduce the test statistic

$$\mathrm{LM}(r_0) = \mathrm{tr}\{\hat{\rho}^* \hat{M}_{vv\cdot z} \hat{\rho}^{*\prime} (\tilde{\alpha}'_\perp \tilde{\Omega} \tilde{\alpha}_\perp)^{-1}\}, \tag{3.3}$$

where $\hat{\rho}^*$ is the LS estimator of ρ^* from (3.2) and

$$\hat{M}_{vv\cdot z} = \sum_{t=p+1}^{T} \hat{v}_{t-1}\hat{v}'_{t-1} - \sum_{t=p+1}^{T} \hat{v}_{t-1}\hat{z}'_t \left(\sum_{t=p+1}^{T} \hat{z}_t \hat{z}'_t\right)^{-1} \sum_{t=p+1}^{T} \hat{z}_t \hat{v}'_{t-1}$$

with $\hat{z}_t = [\hat{u}'_{t-1} \colon \Delta\hat{x}'_{t-1} \colon \cdots \colon \Delta\hat{x}'_{t-p+1}]'$. S&L also discussed an asymptotically equivalent variant of the preceding test statistic obtained by deleting the regressor $\hat{u}_{t-1}$ from (3.2). We shall not consider this modification because it was found to give very similar results in small samples.

The LR-type statistic based on (3.1) is obtained in the usual way by solving the generalized eigenvalue problem

$\det(\bar{\Pi}\hat{M}_T\bar{\Pi}' - \lambda\bar{\Omega}) = 0$, where $\bar{\Pi}$ is the LS estimator of Π obtained from (3.1), $\bar{\Omega}$ is the corresponding residual covariance matrix, and

$$\hat{M}_T = \sum_{t=p+1}^{T} \hat{x}_{t-1}\hat{x}_{t-1}' - \sum_{t=p+1}^{T} \hat{x}_{t-1}\Delta\hat{X}_{t-1}' \times \left(\sum_{t=p+1}^{T} \Delta\hat{X}_{t-1}\Delta\hat{X}_{t-1}'\right)^{-1} \sum_{t=p+1}^{T} \Delta\hat{X}_{t-1}\hat{x}_{t-1}'$$

with $\Delta\hat{X}_{t-1} = [\Delta\hat{x}_{t-1}' : \cdots : \Delta\hat{x}_{t-p+1}']'$. Denoting the resulting eigenvalues by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$, the LR-type statistic becomes

$$\mathrm{LR}(r_0) = \sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j). \tag{3.4}$$

Now we can state the following theorem, where $\mathbf{B}(s)$ is an $(n-r_0)$-dimensional standard Brownian motion.

Theorem 3.1. If $H_0(r_0)$ in (1.12) is true,

$$\mathrm{LM}(r_0), \mathrm{LR}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{B}_*(s)\, d\mathbf{B}_*(s)'\right)' \times \left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\, ds\right)^{-1} \left(\int_0^1 \mathbf{B}_*(s)\, d\mathbf{B}_*(s)'\right)\right\},$$

where $\mathbf{B}_*(s) = \mathbf{B}(s) - s\mathbf{B}(1)$ is an $(n-r_0)$-dimensional Brownian bridge and $d\mathbf{B}_*(s) = d\mathbf{B}(s) - ds\mathbf{B}(1)$ is as in theorem 2 of S&L and theorem 5.1 of L&S. Note that $\int_0^1 \mathbf{B}_*(s)\ d\mathbf{B}_*(s)'$ abbreviates $\int_0^1 \mathbf{B}(s)\ d\mathbf{B}(s)' - \mathbf{B}(1)\int_0^1 s d\mathbf{B}(s)' - \int_0^1 \mathbf{B}(s)\ ds\mathbf{B}(1)' + \frac{1}{2}\mathbf{B}(1)\mathbf{B}(1)'$.

The limiting distribution obtained in Theorem 3.1 is free of unknown nuisance parameters and actually the same as the one obtained by S&L and L&S in a model without any dummy variables. Critical values are given in table 1 of L&S. Thus, in our framework, including step dummies and impulse dummies in the model and estimating their coefficients has no effect on the limiting distribution of the cointegration tests. This is very convenient and contrasts with the LR tests proposed by Johansen and Nielsen (1993). They include dummy variables in the ECM for y_t and show that in this case the asymptotic null distribution depends on the break point. Hence, a new set of critical values is required for each break point. Our theorem extends previous results of S&L who noticed that the limiting distribution of the corresponding tests for models without dummy variables is not affected by the limiting properties of the (GLS) estimator of the mean parameter μ_0.

The preceding discussion also suggests that if the a priori restriction $\mu_1 = 0$ is employed in (1.1) and the GLS estimation of Section 2 as well as the preceding test procedure are modified accordingly, the limiting distribution of the resulting test statistic is the same as in a model without any deterministic terms—that is, the limiting distribution is obtained by replacing the Brownian bridge $\mathbf{B}_*(s)$ in Theorem 3.1 by the Brownian motion $\mathbf{B}(s)$. This situation was also studied by S&L and, from the proofs given in the Appendix and in that article, it can be seen that the preceding conclusion actually holds. The same result can also be obtained by replacing the GLS estimator considered in this article by one proposed by Saikkonen and Luukkonen (1997). This GLS estimation is similar to the one developed in Section 2 except that the estimator $\tilde{A}(L)$ is obtained by ignoring the cointegration structure of (1.11) and applying LS. Note, however, that, to the best of our knowledge, this GLS estimation has not been studied in a model with a time trend.

It may also be worth noting that seasonal dummy variables may be included in the model. They may be treated in the same way as the other dummy variables. It can be shown that including them does not change the asymptotic properties of the tests. Seasonal dummies will be important in the context of the application discussed in Section 5. The small-sample properties of our tests are considered in Section 4.

4. SMALL-SAMPLE COMPARISON OF TESTS

We have compared the properties of the LR and LM tests in a small Monte Carlo experiment. The simulations are based on the following bivariate process x_t from Toda (1994, 1995), which was also used by L&S:

$$x_t = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix} x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{iid}\ N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}\right). \tag{4.1}$$

For $\psi_1 = \psi_2 = 1$, a cointegrating rank of $r = 0$ is obtained. In this case the process consists of two nonstationary components that are independent for $\theta = 0$, whereas they are instantaneously correlated for $\theta \neq 0$. The cointegrating rank is $r = 1$ if $\psi_2 = 1$ and $|\psi_1| < 1$. Furthermore, the process is I(0) with $r = 2$ if both ψ_1 and ψ_2 are less than 1 in absolute value. Because the test results are invariant to the parameter values of the deterministic component, we use $\mu_i = 0$ and $\delta_i = 0$ $(i = 0, 1)$ throughout. In other words, the deterministic part is actually 0. Thereby we can compare the performance of our tests with other tests that do not allow for shifts.

Samples of sizes 100 and 200 plus 50 presample values starting with an initial value of 0 were generated. Thus, we are not using zero initial values in the actual samples in contrast to the theoretical analysis of the previous sections. By considering nonzero initial values, we are able to check the robustness of the theoretical results to violations of the initial value assumptions. The number of replications is 1,000. Rejection frequencies of the tests are given in Tables 1 and 2. They are based on asymptotic critical values for a test level of 5%. The rejection frequencies are not corrected for the actual small-sample sizes because such corrections will not be available in practice. Comparing the power of tests that have unknown size is of limited value from a practical point of view. Therefore, a minimal requirement for a test is that it observes the selected significance level at least approximately.

Table 1. Relative Rejection Frequencies of Test Statistics for DGP (4.1) With Cointegrating Rank $r = 0$ ($\psi_1 = \psi_2 = 1$), $\theta = 0$, $T = 100$, Nominal Significance Level .05

Assumed break point	Test statistic	Rank under H_0: $r_0 = 0$	$r_0 = 1$
None	$LR_{Johansen}$	.060	.005
	LR_{ta}	.052	.002
	LM_{ta}	.033	.003
$T_0 = T_1 = 25$	LR	.058	.009
	LM	.032	.012
$T_0 = T_1 = 50$	LR	.054	.009
	LM	.033	.008
$T_0 = T_1 = 75$	LR	.049	.007
	LM	.036	.012

For a given set of parameter values and a given sample size, the results for the test statistics are based on the same generated time series. Hence, the entries in the tables are not independent but can be compared directly. Still, for judging the results, it may be worth recalling that the standard error of an estimator of a true rejection probability P based on 1,000 replications of the experiment is $s_P = \sqrt{P(1-P)/1{,}000}$ so that, for example, $s_{.05} = .007$. It is also important to note that in the simulations the tests were not performed sequentially. Thus, the results for testing $H_0(1)$: $\mathrm{rk}(\Pi) = 1$ are not conditioned on the outcome of the test of $H_0(0)$: $\mathrm{rk}(\Pi) = 0$.

In the tables the performance of the tests that allow for a structural shift is compared to tests that allow for a linear trend only and not for shifts in the deterministic term. Note that in the literature on testing for a unit root, which may be regarded as a special case of testing for cointegration, it is a well-established fact that ignoring a structural change in the DGP can lead to substantial distortions (e.g., see Perron 1989). Therefore, if a structural shift is suspected, one may want to allow for it. Then the question arises, "What are the implications of including a shift term if no shift has occurred?" The answer to this question can be explored by including tests that do not allow for a structural shift in a comparison. The tests denoted as $\mathrm{LR}_{\mathrm{Johansen}}$, LR_{ta}, and LM_{ta} are the appropriate LR and LM tests proposed by Johansen (1995) and S&L for the latter situation. In Table 1, the true cointegrating rank is 0. Thus, the results for testing $H_0(0)$: $\mathrm{rk}(\Pi) = 0$ give an indication of the actual sizes of the tests for a nominal size of 5%. It is seen that the location of the break point (T_0, T_1) does not matter much for the size of the tests in samples with $T = 100$. The LM tests are generally a bit more conservative than the LR tests. This also holds for the tests that do not allow for a shift. In fact, whether or not a shift is accommodated does not matter much for the sizes of the tests. All tests tend to be very conservative if the rank is overstated in the null hypothesis (see $r_0 = 1$ in Table 1).

In Table 2 we focus on the case in which a possible break occurs after three quarters of the sample. The location of the break point turned out to have little impact on the test performance. A break point closer to the end than to the beginning of the sample as in the table is relevant in the context of the empirical application of Section 5 and is therefore considered here. A test of $r_0 = 0$ shows the power of the tests when $|\psi_1| < 1$. As expected, the LR tests have slightly larger power than the corresponding LM tests because they are less conservative. It is also seen that there is some gain in power if the a priori information is used that there is actually no shift in the DGP. In other words, in terms of power $\mathrm{LR}_{\mathrm{Johansen}}$ and LR_{ta} outperform the LR and LM tests that allow for a shift. Notice, however, that LM_{ta} is less powerful than LR. Thus, it pays to use LR-type tests rather than LM tests. Moreover, it is useful to take into account the prior information of no shift if such information is available. In our particular DGP, the power is generally higher if there is more residual correlation ($\theta = .8$). Both tests that allow for a shift tend to be conservative in checking the null hypothesis $r_0 = 1$.

We have also performed the simulations with samples of the size $T = 200$. As expected, the power increases. The tests still tend to be conservative for $r_0 = 1$, however. We do not show the results because the marginal information content does not justify the additional space needed. In Section

Table 2. Relative Rejection Frequencies of Test Statistics for DGP (4.1) With Cointegrating Rank $r = 0$ ($\psi_1 = 1$) or 1 ($\psi_1 < 1$), $\psi_2 = 1$, Sample Size $T = 100$, Break Point $T_0 = T_1 = 75$, Nominal Significance Level .05

Statistic	$\psi_1 = 1.0$: $r_0 = 0^*$	$r_0 = 1^*$	$\psi_1 = .95$: $r_0 = 0^*$	$r_0 = 1^*$	$\psi_1 = .9$: $r_0 = 0^*$	$r_0 = 1^*$	$\psi_1 = .8$: $r_0 = 0^*$	$r_0 = 1^*$
				$\theta = 0$				
$LR_{Johansen}$	.060	.005	.068	.007	.098	.008	.292	.024
LR_{ta}	.052	.002	.065	.007	.104	.030	.313	.047
LM_{ta}	.033	.003	.044	.008	.075	.018	.253	.042
LR	.049	.007	.056	.015	.092	.019	.289	.044
LM	.036	.012	.045	.012	.063	.030	.218	.047
				$\theta = .8$				
$LR_{Johansen}$	.060	.005	.155	.012	.421	.040	.940	.065
LR_{ta}	.052	.002	.147	.011	.396	.017	.828	.031
LM_{ta}	.033	.003	.113	.006	.326	.016	.778	.030
LR	.049	.007	.125	.014	.355	.017	.793	.031
LM	.036	.012	.056	.018	.294	.020	.735	.029

* Rank specified in the null hypothesis.

5, we will apply our new tests to analyze the cointegrating ranks of actual economic systems.

5. GERMAN MONEY-DEMAND SYSTEMS

Based on a single equation analysis of Wolters, Teräsvirta, and Lütkepohl (1998), Lütkepohl and Wolters (1998) constructed a small macroeconomic model to investigate the channels of German monetary policy. They built a vector ECM for M3, GNP, an inflation rate, an interest-rate spread variable, and import-price inflation. In a demand relation for M3, GNP is a proxy for the transactions volume, the inflation rate and the interest-rate spread are opportunity-cost variables, and the import-price inflation is included as a measure for the real exchange rate to account for the openness of the German economy. The variable M3 is used to measure the money stock because it used to be the target variable of the Deutsche Bundesbank in executing its monetary policy. The interest-rate spread and the import-price inflation variables turned out to be stationary, whereas the other three variables were found to be I(1) in the aforementioned studies. Therefore, we shall first focus on M3, GNP, and the inflation rate in the following and investigate the number of cointegration relations among these variables. For illustrative purposes, we will then add the interest-rate spread to the system to check whether a further cointegration relation is found by the tests.

We use quarterly, seasonally unadjusted data for the period 1975 to 1996 as used by Lütkepohl and Wolters (1998). The initial period was chosen because the Bundesbank started its policy of monetary targeting in 1975. Specifically the following variables are used: m_t represents the logarithm of real M3, gnp_t is the logarithm of real GNP, r_t is the difference between the average bond rate and the own rate of M3, and p_t is the logarithm of the GNP deflator; hence, Δp_t is the inflation rate that will be used here. In other words, the two systems of variables considered in the following cointegration analyses are $y_t = (m_t, gnp_t, \Delta p_t)'$ and $y_t = (m_t, gnp_t, \Delta p_t, r_t)'$. The data sources are given in Appendix B. The variables are plotted in Figure 1. Obviously, m_t and gnp_t have clear shifts in the third quarter of 1990. These shifts are due to the German reunification. Note that, although the political reunification took place in October 1990, the monetary unification occurred already on July 1, 1990. Since then, all variables refer to the unified Germany and, hence, the shift in the third quarter of 1990 is quite natural.

If the full sample is used in the following, the structural shifts in m_t and gnp_t in 1990 have to be taken into account. Therefore, Lütkepohl and Wolters (1998) and Wolters et al. (1998) included a shift dummy $d_{1t} = 0$ until the second quarter of 1990 and $d_{1t} = 1$ afterwards. They also included an impulse dummy $d_{0t} = 1$ in 1990(3) and 0 elsewhere. We will use these dummy variables in our analysis as well. Because m_t and gnp_t potentially have a deterministic trend term, we also include a linear trend in the models. In addition we also need seasonal dummy variables for some of the series because our data are not seasonally adjusted and have quite pronounced seasonal components (see Fig. 1). In

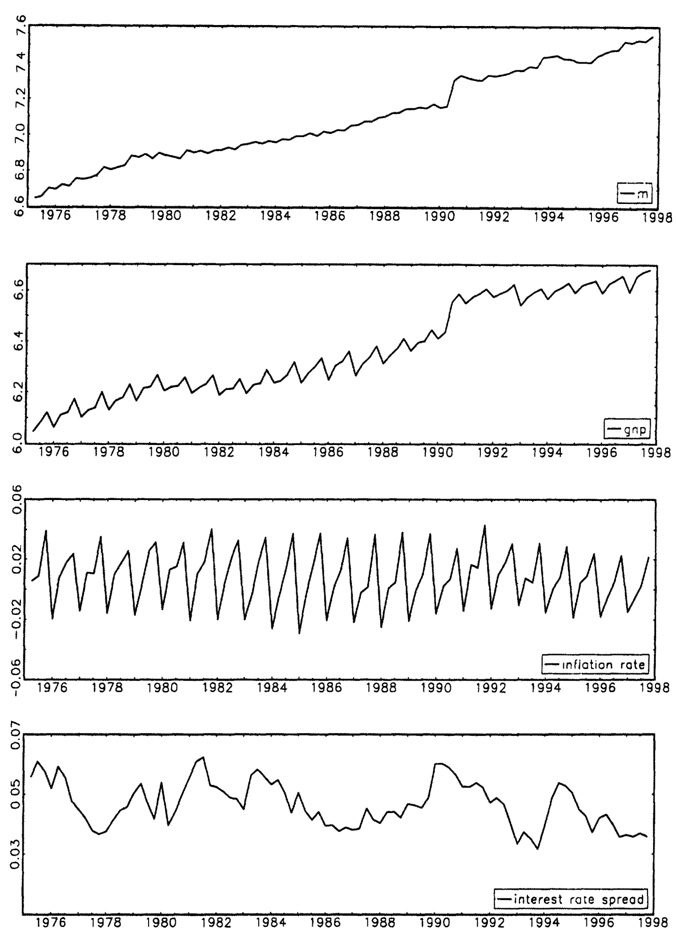

Figure 1. Time Series Analyzed.

the following, we will estimate the parameters associated with the seasonal dummy variables by the GLS procedure described in Section 2 for the other parameters of the deterministic part of the model. Then y_t is adjusted for all deterministic terms including the seasonal terms to get $\hat{x}_t$.

Before we perform a cointegration analysis, some comments on the integration properties of the individual variables are in order. They were investigated by Wolters et al. (1998), who applied augmented Dickey–Fuller (ADF) tests using only preunification data until 1989(4). Based on the reduced observation period, m_t, gnp_t, and Δp_t were found to be I(1), whereas there was evidence of r_t being I(0). For the extended sampling period until 1996(4), we have confirmed these results for Δp_t and r_t. Conventional ADF tests may be used for these variables because they do not have a shift due to the unification. On the other hand, the shift in m_t and gnp_t has to be taken into account in unit-root tests. Therefore, we use our new tests for this purpose by checking H_0: rk$(\Pi) = 0$ against H_1: rk$(\Pi) = 1$ for the individual series. The results are shown in Table 3. In the table we also give results for the preunification period under the assumption that no shift has occurred in that period. Two different autoregressive orders are considered, and it is seen that the results are robust with respect to the order. In none of the cases can the unit root be rejected. Thereby the preunification results are also confirmed for these series.

Table 3. Unit-Root Tests for German M3 and GNP Data

Variable	Sample period (deterministic terms)	Autoregressive order	LR*	LM*
m_t	1975(3)–1990(2) (T = 60)	p = 2	1.07	1.04
	1975(3)–1990(2) (T = 60) (seas. dummies, lin. trend)	p = 4	.95	.95
	1975(3)–1996(4) (T = 86)	p = 2	2.45	2.44
	1975(3)–1996(4) (T = 86) (seas. dum., lin. trend, d_{0t}, d_{1t})	p = 4	2.78	2.68
gnp_t	1975(3)–1990(2) (T = 60)	p = 2	1.42	1.44
	1975(3)–1990(2) (T = 60) (seas. dummies, lin. trend)	p = 4	1.06	1.07
	1975(3)–1996(4) (T = 86)	p = 2	2.03	2.08
	1975(3)–1996(4) (T = 86) (seas. dum., lin. trend, d_{0t}, d_{1t})	p = 4	1.13	1.10

* Critical values: 6.83 (5%), 5.43 (10%) (from table 1 of L&S).

We have also checked for seasonal unit roots using only preunification data. There is some evidence for a possible root at the semiannual frequency in gnp_t, as well as for roots at the annual frequencies in gnp_t and Δp_t. On the other hand, no seasonal unit roots are found in m_t. Of course, r_t does not have a seasonal component (see also Fig. 1). Thus, the seasonal structure of the series seems to be quite different. Still, the possibility of seasonal cointegration between gnp_t and Δp_t cannot be excluded on the basis of these results. Nevertheless, we ignore this potential problem in the following analysis, which is meant to illustrate our new tests. Clearly, we are far from having a suitable framework for analyzing seasonal cointegration in the presence of structural shifts [see Johansen and Schaumburg (1999) for a discussion of seasonal cointegration].

In Table 4 the results of various cointegration tests for the three-dimensional system are provided. They are based on models of order $p = 2$, which was also used by Wolters et al. (1998). This order is recommended by the Hannan–Quinn criterion (HQ). Seasonal dummies and linear-trend terms, as well as d_{0t} and d_{1t}, are included as deterministic terms in all models for the full sampling period. In addition to tests for the full period, we also give results for the preunification period using data up to 1990(2) only. In these tests the dummies d_{it} $(i = 0, 1)$ are not included. Obviously, one cointegration relation is found regardless of the observation period. Moreover, for common significance levels both versions of the tests clearly reach the same conclusion regarding the cointegrating rank. Still it is pleasing that the new tests enable us to use the full-sample information.

Table 4. Cointegration Tests for Three-dimensional System (m_t, gnp_t, Δp_t)

H_0	Critical values* 90%	Critical values* 95%	1975(3)–1990(2) (T = 60) LR	1975(3)–1990(2) (T = 60) LM	1975(3)–1996(4) (T = 86) LR	1975(3)–1996(4) (T = 86) LM
$r_0 = 0$	25.90	28.47	38.94	29.63	52.71	42.55
$r_0 = 1$	13.89	15.92	2.39	2.36	5.56	5.73
$r_0 = 2$	5.43	6.83	.63	.33	1.26	1.57

* From table 1 of L&S.

Table 5. Cointegration Tests for Four-Dimensional System (m_t, gnp_t, Δp_t, r_t)

H_0	Critical values* 90%	Critical values* 95%	1975(3)–1990(2) (T = 60) LR	1975(3)–1990(2) (T = 60) LM	1975(3)–1996(4) (T = 86) LR	1975(3)–1996(4) (T = 86) LM
$r_0 = 0$	42.03	45.13	55.21	43.13	78.28	65.33
$r_0 = 1$	25.90	28.47	15.62	14.18	30.31	28.21
$r_0 = 2$	13.89	15.92	3.00	2.99	4.94	5.05
$r_0 = 3$	5.43	6.83	.74	.21	1.34	2.21

* From table 1 of L&S.

A similar analysis was also performed for the four-dimensional system $y_t = (m_t, gnp_t, \Delta p_t, r_t)'$. The results are given in Table 5. The VAR order of 2 is chosen because it was also used for the three-dimensional system, although HQ now recommends a VAR order of 1. Because r_t was found to be stationary in the univariate analysis, one would expect to find a cointegrating rank of 2 for the present system, the second "cointegration relation" consisting of the stationary variable only. It turns out, however, that tests based on the preunification period clearly indicate a cointegrating rank of 1. It may be worth noting that the same result is obtained with the corresponding Johansen LR test. Extending the sample period to 1996, however, and including the shift dummy variables d_{0t} and d_{1t} in the system, the expected cointegrating rank of 2 is found. Thus, the result for the shorter sampling period may be due to reduced power, and in this case it clearly pays to take full advantage of the available data. Interestingly, in this case the LM version of our test strictly speaking rejects one cointegration relation only at the 10% level although the test value is close to the 5% critical value. In contrast, the LR version rejects at the 5% level. This result reflects the reduced power of the LM tests found in the simulations of Section 4.

6. CONCLUSIONS

In this study we have proposed and applied tests for the cointegrating rank of a system of variables in the presence of structural shifts. Under the assumption that the break point is known a priori, we suggest estimating the deterministic part of the DGP first, subtracting the estimated deterministic part from the original series, and then performing standard systems cointegration tests for the adjusted series. We have considered LR- and LM-type tests in this context and find that they have asymptotic null distributions that are tabulated elsewhere in the literature and do not depend on the break point. Hence, the tests are conveniently applied without the need for simulating new critical values. In a small simulation study, it is found that the LR versions of the tests tend to have more power than the LM versions. Therefore, the use of the LR-type tests is proposed. For illustrative purposes, the tests are applied to two systems of German macroeconomic variables that may be thought of as money-demand systems. It is found that taking into account the level shift in some of the series is necessary for proper inference regarding the cointegrating rank of the system.

ACKNOWLEDGMENTS

We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract no. ERBFMRXCT980213) for financial support. We thank Carsten Trenkler and Kirstin Hubrich for helping with the computations. Comments by three anonymous referees are gratefully acknowledged. Part of this research was done while the first author was visiting the Humboldt University in Berlin.

APPENDIX A: PROOFS

In the following proofs we will skip some of the details to save space. They can be found in the discussion-paper version of the article, which is available on request from the second author.

Proof of Lemma 2.1

It is not difficult to see that assuming μ_i and δ_i to be 0 $(i = 0, 1)$ does not imply a loss of generality so that in (1.11) the true values of ν, τ, θ, and γ_{ij} are also all 0 and $y_t = x_t$ can be assumed. Define

$$\begin{aligned} X_t &= [y'_{t-1}\colon t-1\colon d_{1,t-1}]' \\ Z_{1t} &= [1\colon \Delta y'_{t-1}\colon \cdots \colon \Delta y'_{t-p+1}]' \\ Z_{2t} &= [d_{0t}\colon \cdots \colon d_{0,t-p}\colon \Delta d_{1t}\colon \cdots \colon \Delta d_{1,t-p+1}]' \end{aligned}$$

and set $Z_t = [Z'_{1t}\colon Z'_{2t}]'$. Then we can write Equation (1.11) as

$$\Delta y_t = \alpha\psi' X_t + \Phi Z_t + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \quad \text{(A.1)}$$

where $\psi' = [\beta'\colon -\tau\colon -\theta]$, $\Phi = [\Phi_1\colon \Phi_2]$ with $\Phi_1 = [\nu\colon \Gamma_1 : \cdots : \Gamma_{p-1}]$, and $\Phi_2 = [\gamma_{00} : \cdots : \gamma_{0p}\colon \gamma_{10} : \cdots : \gamma_{1,p-1}]$. The RR regression estimators of α, ψ, and Ω can be obtained as follows. Define

$$S_{ij} = N^{-1} \sum_{t=p+1}^{T} R_{it} R'_{jt}, \qquad i, j = 0, 1,$$

where $N = T - p$ and R_{0t} and R_{1t} are the LS residuals obtained by regressing Δy_t and X_t on Z_t, respectively. As is well known, the RR regression estimator of ψ is based on the eigenvectors corresponding to the r largest eigenvalues of the determinantal equation

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \qquad \text{(A.2)}$$

(e.g., Johansen 1995, chap. 6, or Lütkepohl 1991, appendix A.14). When the RR regression estimator of ψ is available, those of α and Ω are obtained by replacing ψ by its estimator in the formulas $S_{01}\psi(\psi' S_{11}\psi)^{-1}$ and $S_{00} - S_{01}\psi(\psi' S_{11}\psi)^{-1}\psi' S_{10}$, respectively. Recall that we have assumed that all impulse dummies in (1.11) are linearly independent so that S_{00} is nonsingular and the preceding estimators are well defined.

In the same way as in the proof of lemma 13.1 of Johansen (1995), we now transform Equation (A.2). To this end, we define the matrix

$$A_T = \begin{bmatrix} \beta & T^{-1/2}\beta_\perp(\beta'_\perp\beta_\perp)^{-1} & 0 & 0 \\ 0 & 0 & T^{-1} & 0 \\ 0 & 0 & 0 & T^{1/2}/(T-T_1)^{1/2} \end{bmatrix}.$$

Pre- and postmultiplying (A.2) by A'_T and A_T, respectively, gives

$$\det(\lambda A'_T S_{11} A_T - A'_T S_{10} S_{00}^{-1} S_{01} A_T) = 0. \qquad \text{(A.3)}$$

The eigenvalues of (A.2) and (A.3) are identical and the eigenvectors of (A.3) are obtained from those of (A.2) by premultiplying by A_T^{-1}. The next step is to study the weak limit of (A.3). For ease of exposition and without loss of generality, we shall assume that the initial values of x_t are such that $\beta' x_t$ and Δx_t are stationary.

Next note that

$$S_{00} = \tilde{S}_{00} + o_p(1) \quad \text{and} \quad S_{01}A_T = \tilde{S}_{01}A_T + o_p(1), \quad \text{(A.4)}$$

where $\tilde{S}_{00}$ and $\tilde{S}_{01}$ are analogs of S_{00} and S_{01}, respectively, obtained by omitting Z_{2t} from (A.1) (i.e., $Z_t = Z_{1t}$). The proof of (A.4) is a straightforward consequence of well-known limit theorems and the fact that Z_{2t} is a vector of impulse dummies.

From (A.4) we can conclude that the asymptotic behavior of the latter matrix in (A.3) is similar to that in the conventional model without any dummies. Thus, in the same way as Johansen (1995, pp. 158, 180), we find that

$$A'_T S_{10} S_{00}^{-1} S_{01} A_T \xrightarrow{p} \begin{bmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{(A.5)}$$

where Σ_{00} and $\Sigma_{0\beta}$ are conditional covariance matrices defined by

$$\operatorname{cov}\left(\begin{bmatrix} \Delta x_t \\ \beta' x_t \end{bmatrix} | \Delta x_{t-1}, \ldots, \Delta x_{t-p+1} \right) = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}.$$

Next we have to consider the asymptotic behavior of the matrix $A'_T S_{11} A_T$. For this purpose we note that

$$A'_T X_t = \begin{bmatrix} \beta' y_{t-1} \\ T^{-1/2}(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp y_{t-1} \\ T^{-1}(t-1) \\ \left(\frac{T}{T-T_1}\right)^{1/2} d_{1,t-1} \end{bmatrix}, \qquad \text{(A.6)}$$

where $y_t = x_t$ can be assumed. It appears convenient to analyze separately the cases in which $a_1 < 1$ and $a_1 = 1$ in (1.8).

$a_1 < 1$. If $a_1 < 1$, we can proceed in the same way as in the case of (A.4) and show that $A'_T S_{11} A_T = A'_T \tilde{S}_{11} A_T + o_p(1)$ with $\tilde{S}_{11}$ defined by omitting Z_{2t} in the same way as in $\tilde{S}_{00}$ and $\tilde{S}_{01}$. Because we may assume that $y_t = x_t$, the weak limit of $A'_T \tilde{S}_{11} A_T$ can be obtained by arguments similar to those used in the conventional case (see Johansen 1995, pp. 158, 180). It is first easy to check that the matrix $A'_T \tilde{S}_{11} A_T$ is asymptotically block diagonal with the $(r \times r)$ block in the upper left corner converging in probability to $\Sigma_{\beta\beta}$. To obtain the weak limit of the lower right corner, let $\mathbf{W}(s)$ be an n-dimensional Brownian motion with covariance matrix Ω, denote by $\mathrm{I}(\cdot)$ the indicator function, and let $[c]$ be the integer part of c. From the multivariate invariance principle, we then find that

$$\begin{bmatrix} T^{-1/2}\beta_\perp' x_{[Ts]-1} \\ T^{-1}([Ts]-1) \\ \left(\frac{T}{T-T_1}\right)^{1/2} d_{1,[Ts]-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \beta_\perp' C\mathbf{W}(s) \\ s \\ (1-a_1)^{-1/2}\mathrm{I}(s \geq a_1) \end{bmatrix} \overset{\mathrm{def}}{=} \mathbf{G}(s), \quad (A.7)$$

where convergence is in the Skorohod topology of $D[0,1]$. Hence, we can conclude that

$$\lambda A_T' S_{11} A_T \xrightarrow{d} \begin{bmatrix} \lambda\Sigma_{\beta\beta} & 0 \\ 0 & \lambda\int_0^1 \bar{\mathbf{G}}(s)\bar{\mathbf{G}}(s)'\,ds \end{bmatrix}, \quad (A.8)$$

where $\bar{\mathbf{G}}(s) = \mathbf{G}(s) - \int_0^1 \mathbf{G}(u)\,du$ (cf. Johansen 1995, pp. 158, 180).

Now, let $\tilde{\psi} = [\tilde{\beta}_\xi', -\tilde{\tau}, -\tilde{\theta}]'$ be the (normalized) RR regression estimator of ψ described previous to (A.2). In the same way as Johansen (1995, p. 180), we can then use (A.5) and (A.8) to conclude that

$$A_T^{-1}\tilde{\psi} = \begin{bmatrix} I_r \\ T^{1/2}\beta_\perp'\tilde{\beta}_\xi \\ -T\tilde{\tau}' \\ -\left(\frac{T-T_1}{T}\right)^{1/2}\tilde{\theta}' \end{bmatrix} = \begin{bmatrix} I_r \\ o_p(1) \\ o_p(1) \\ o_p(1) \end{bmatrix}, \quad (A.9)$$

where the first equality follows from the definitions of A_T and $\tilde{\psi}$. This shows that the estimators $\tilde{\beta}_\xi, \tilde{\tau}$, and $\tilde{\theta}$ are consistent of orders $o_p(T^{-1/2}), o_p(T^{-1})$, and $o_p(1)$, respectively. (Recall that the true values of τ and θ are 0.) From these results, it is further straightforward to show that $\tilde{\alpha}_\xi; \tilde{\Gamma}_j$, and $\tilde{\Omega}$ are also consistent (cf. Johansen 1995, p. 181). Before strengthening the preceding consistency results to the required form, we shall briefly discuss similar intermediate results in the case $a_1 = 1$.

$a_1 = 1$. If $a_1 = 1$, the previous derivation of (A.8) fails because the limit in (A.7) is not defined. Using arguments similar to those used to prove (A.4) and (A.8), however, it can be shown that instead of (A.8) we now have

$$\lambda A_T' S_{11} A_T \xrightarrow{d} \begin{bmatrix} \lambda\Sigma_{\beta\beta} & 0 & 0 \\ 0 & \lambda\int_0^1 \bar{\mathbf{G}}_1(s)\bar{\mathbf{G}}_1(s)'\,ds & 0 \\ 0 & 0 & \lambda(1-c_1^2) \end{bmatrix}, \quad (A.10)$$

where $0 < c_1 < 1$ and $\bar{\mathbf{G}}_1(s) = \mathbf{G}_1(s) - \int_0^1 \mathbf{G}_1(u)\,du$ with $\mathbf{G}_1(s)((n-r+1)\times 1)$ defined by the first $n-r+1$ components of $\mathbf{G}(s)$. Thus, because we still have (A.5), the only difference between the present case ($a_1 = 1$) and the previous one ($a_1 < 1$) is that the weak limit of $\lambda A_T' S_{11} A_T$ differs from its previous counterpart obtained in (A.8). This, however, does not affect the arguments used to derive (A.9), which therefore hold in the present context as well. This implies that $\tilde{\beta}_\xi$ and $\tilde{\tau}$ are consistent of the same orders as previously but, because $T_1/T \to 1$, nothing can be concluded about the consistency of $\tilde{\theta}$. Fortunately, however, the consistency of $\tilde{\alpha}_\xi, \tilde{\Gamma}_j$, and $\tilde{\Omega}$ can still be proved in the same way as before. To see this, note that $\tilde{\alpha}_\xi$ and $\tilde{\Gamma}_j$ can be obtained by LS from the auxiliary regression model

$$\Delta y_t = \alpha\tilde{\psi}_T' A_T' X_t + \nu + \sum_{j=1}^{p-1}\Gamma_i \Delta y_{t-j} + \Phi_2 Z_{2t} + \text{error}, \quad t = p+1,\ldots,T, \quad (A.11)$$

where $\tilde{\psi}_T = A_T^{-1}\tilde{\psi}$. Because $A_T^{-1}\psi = [I_r\colon 0]'$, $\tilde{\psi}_T$ is a consistent estimator of $A_T^{-1}\psi$ [see (A.9)]. This consistency result and results about the second sample moments of the variables in the auxiliary regression model (A.11), used to show that (A.5), (A.8), and (A.10) hold, can be used to prove the consistency of the estimators $\tilde{\alpha}_\xi$ and $\tilde{\Gamma}_j$ (and also $\tilde{\nu}$). After this, the consistency of $\tilde{\Omega}$ can be proved in a standard fashion.

To complete the proof, we have to establish the stated orders of consistency. Define

$$U_T = \begin{bmatrix} T\beta_\perp'\tilde{\beta}_\xi \\ T^{3/2}\tilde{\tau}' \\ (T-T_1)^{1/2}\tilde{\theta}' \end{bmatrix} \; ((n-r+2)\times r).$$

First, we wish to show that $U_T = O_p(1)$, which implies that $T(\tilde{\beta}_\xi - \beta) = O_p(1)$. Define the $((n+2)\times(n-r+2))$ matrix

$$B_T = \begin{bmatrix} \beta_\perp(\beta_\perp'\beta_\perp)^{-1} & 0 & 0 \\ 0 & T^{-1/2} & 0 \\ 0 & 0 & \frac{T}{(T-T_1)^{1/2}} \end{bmatrix}$$

and notice that the matrix A_T can be written as $A_T = [b\colon T^{-1/2}B_T]$ with $b = [\beta' \;\; 0 \;\; 0]'$. Because U_T is formed by the last $n-r+2$ rows of $T^{1/2}A_T^{-1}(\tilde{\psi} - \psi)$, it follows from the preceding definitions that (cf. Johansen 1995, p. 179) $\tilde{\psi} - \psi = T^{-1}B_T U_T$. Using the first-order conditions for $\tilde{\psi}$, the result $U_T = O_p(1)$ and, hence, $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$ can now be established. Details, based on results of second sample moments already used in previous steps of the proof, are very similar to those of Johansen (1995, p. 182). Moreover, because $U_T = O_p(1)$ implies that $\tilde{\psi}_T$ in (A.11) is a consistent estimator of $A_T^{-1}\psi$ of order $O_p(T^{-1/2})$, the remaining orders of consistency are straightforward to establish by considering the LS estimators obtained from (A.11). This completes the proof of Lemma 2.1.

Remark A.1. (a) From the proof it is also possible to derive the limiting distributions of $\tilde{\alpha}_\xi$ and $\tilde{\beta}_\xi$ and thereby similar results for other normalizations. The limiting distribution of $\tilde{\beta}_\xi$ is mixed normal, but its explicit form differs for the cases $a_1 < 1$ and $a_1 = 1$ [cf. (A.8) and (A.10)].

(b) From the proof it can also be seen that $\tilde{\tau} - \tau = O_p(T^{-3/2})$ and $\tilde{\theta} - \theta = O_p((T - T_1)^{-1/2})$. Because $\tau = \beta'\mu_1$ and $\theta = \beta'\delta_1$, these results may be viewed as RR analogs of (2.9) and (2.7).

Proof of Theorem 2.1

Because all relevant quantities will be invariant to normalizations of $\tilde{\alpha}$ and $\tilde{\beta}$, we can assume that some kind of normalization has been imposed so that, by Lemma 2.1, we can also assume that $\tilde{\alpha} = \alpha + O_p(T^{-1/2})$ and $\tilde{\beta} = \beta + O_p(T^{-1})$.

Using the definitions, we first note that explicit forms of the variables $\tilde{H}_{it}$ and $\tilde{K}_{it}$ $(i = 0, 1)$ in (2.3) are given by

$$\tilde{H}_{0t} = \begin{cases} I_n, & t = 1 \\ I_n - \sum_{j=1}^{t-1} \tilde{A}_j, & t = 2, \ldots, p \\ -\tilde{\alpha}\tilde{\beta}', & t = p+1, \ldots, T \end{cases},$$

$$\tilde{H}_{1t} = \begin{cases} I_n, & t = 1 \\ tI_n - \sum_{j=1}^{t-1} (t-j)\tilde{A}_j, & t = 2, \ldots, p \\ \tilde{\Psi} - (t-1)\tilde{\alpha}\tilde{\beta}', & t = p+1, \ldots, T \end{cases},$$

$$\tilde{K}_{0t} = \begin{cases} 0, & t < T_0 \\ I_n, & t = T_0 \\ -\tilde{A}_j, & t = T_0 + j, \quad j = 1, \ldots, p \\ 0, & t = T_0 + p + 1, \ldots, T \end{cases},$$

and

$$\tilde{K}_{1t} = \begin{cases} 0, & t < T_1 \\ I_n, & t = T_1 \\ I_n - \sum_{j=1}^{t-T_1} \tilde{A}_j, & t = T_1 + 1, \ldots, T_1 + p - 1 \\ -\tilde{\alpha}\tilde{\beta}', & t = T_1 + p, \ldots, T \end{cases}.$$

The proof is based on ideas similar to those used in the proof of theorem 1 of S&L.

Define

$$\underline{\gamma}_1 = \begin{bmatrix} \underline{\gamma}_{11} \\ \underline{\gamma}_{12} \\ \underline{\gamma}_{13} \end{bmatrix} = \begin{bmatrix} \delta_0 \\ \beta'_\perp \delta_1 \\ \beta'_\perp \mu_0 \end{bmatrix}$$

and

$$\underline{\gamma}_2 = \begin{bmatrix} \underline{\gamma}_{21} \\ \underline{\gamma}_{22} \\ \underline{\gamma}_{23} \\ \underline{\gamma}_{24} \end{bmatrix} = \begin{bmatrix} \tilde{\beta}'\delta_1 \\ \tilde{\beta}'\mu_0 \\ \tilde{\beta}'\mu_1 \\ \tilde{\beta}'_\perp \mu_1 \end{bmatrix}.$$

The idea is to first obtain asymptotic properties of the LS estimator of the "parameters" $\underline{\gamma}_1$ and $\underline{\gamma}_2$ and then use theorem 1 of S&L to derive the stated results. To express (2.3) in terms of $\underline{\gamma}_1$ and $\underline{\gamma}_2$, we transform the variables $\tilde{H}_{it}$ and $\tilde{K}_{it}$ accordingly and define

$$\begin{aligned} \tilde{F}_{11t} &= \tilde{Q}'\tilde{K}_{0t} \\ \tilde{F}_{12t} &= \tilde{Q}'\tilde{K}_{1t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1} \\ \tilde{F}_{13t} &= \tilde{Q}'\tilde{H}_{0t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1} \end{aligned}$$

and

$$\begin{aligned} \tilde{F}_{21t} &= \tilde{Q}'\tilde{K}_{1t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1} \\ \tilde{F}_{22t} &= \tilde{Q}'\tilde{H}_{0t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1} \\ \tilde{F}_{23t} &= \tilde{Q}'\tilde{H}_{1t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1} \\ \tilde{F}_{24t} &= \tilde{Q}'\tilde{H}_{1t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}. \end{aligned}$$

Then, setting $\tilde{F}_{1t} = [\tilde{F}_{11t} : \tilde{F}_{12t} : \tilde{F}_{13t}]$ and $\tilde{F}_{2t} = [\tilde{F}_{21t} : \tilde{F}_{22t} : \tilde{F}_{23t} : \tilde{F}_{24t}]$, we can write (2.3) as

$$\tilde{d}_t := \tilde{Q}'\tilde{A}(L)y_t = \tilde{F}_{1t}\underline{\gamma}_1 + \tilde{F}_{2t}\underline{\gamma}_2 + \eta_t, \qquad t = 1, \ldots, T. \qquad \text{(A.12)}$$

From the definitions, it follows that $\tilde{F}_{1t}$ takes nonzero values only for a fixed number of time indices t.

We shall next study the sums of cross-products between $\tilde{F}_{1t}$ and the error term η_t, which is identical to its counterpart in S&L. Thus,

$$\eta_t = \tilde{Q}'\varepsilon_t - \tilde{Q}'\tilde{\alpha}(\tilde{\beta} - \beta)'x_{t-1} - \tilde{Q}'(\tilde{\alpha} - \alpha)\tilde{\beta}'x_{t-1} - \tilde{Q}'\sum_{j=1}^{p-1}(\tilde{\Gamma}_j - \Gamma_j)\Delta x_{t-j}. \qquad \text{(A.13)}$$

Using this expression, Lemma 2.1, and the previously mentioned property of $\tilde{F}_{1t}$, it can be seen that

$$\sum_{t=1}^{T} \tilde{F}'_{1t}\eta_t = O_p(1). \qquad \text{(A.14)}$$

For $\tilde{F}_{2t}$ the corresponding result is

$$\Upsilon_T^{-1}\sum_{t=1}^{T} \tilde{F}'_{2t}\eta_t = O_p(1), \qquad \text{(A.15)}$$

where $\Upsilon_T^{-1} = \operatorname{diag}[(T - T_1)^{-1/2}I : T^{-1/2}I : T^{-3/2}I : T^{-1/2}I]$ and the partition is conformable to that of $\tilde{F}_{2t}$. To justify (A.15), note first that for $\tilde{F}_{22t}, \tilde{F}_{23t}, \tilde{F}_{24t}$ the result is obtained directly from S&L. When (1.8) holds with $a_1 < 1$, (A.15) can be justified for $\tilde{F}_{21t}$ in the same way as for $\tilde{F}_{22t}$ because $\tilde{F}_{21t} = \tilde{F}_{22,t-T_1+1}$ for $t \geq T_1$ and 0 for $t = 1, \ldots, T_1 - 1$. If (1.8) holds with $a_1 = 1$, the desired result can be obtained by using arguments similar to those that can be used to obtain (A.10).

The next step is to show that the standardized moment matrix of the auxiliary regression model (A.12) converges in probability to a positive definite limit, which in conjunction with (A.14) and (A.15) shows that $\hat{\underline{\gamma}}_1$ and $\hat{\underline{\gamma}}_2$, the LS estimators obtained from (A.12), satisfy

$$\hat{\underline{\gamma}}_1 = \underline{\gamma}_1 + O_p(1) \quad \text{and} \quad \Upsilon_T(\hat{\underline{\gamma}}_2 - \underline{\gamma}_2) = O_p(1). \qquad \text{(A.16)}$$

The derivations needed are based on Lemma 2.1 and arguments similar to those used by S&L.

Because the LS estimators $\hat{\underline{\gamma}}_1$ and $\hat{\underline{\gamma}}_2$ are obtained by replacing the parameters μ_i and δ_i in the definitions of $\underline{\gamma}_1$ and

$\underline{\gamma}_2$ by the LS estimators $\hat{\mu}_i$ and $\hat{\delta}_i$, respectively, it follows from (A.16) that $\hat{\mu}_i = O_p(1)$ and $\hat{\delta}_i = O_p(1)$ $(i = 0, 1)$. From this and Lemma 2.1, one obtains (2.4)–(2.9) in the same way as in the proof of theorem 1 of S&L. For example, (2.7) follows from

$$\begin{aligned}\beta'(\hat{\delta}_1 - \delta_1) &= \tilde{\beta}'(\hat{\delta}_1 - \delta_1) - (\tilde{\beta} - \beta)'(\hat{\delta}_1 - \delta_1) \\ &= O_p((T - T_1)^{-1/2}) + O_p(T^{-1}).\end{aligned}$$

As to the proof of (2.10), it is possible to use the definitions of $\tilde{F}_{1t}$ and $\tilde{F}_{2t}$ and again reduce the problem to that considered by S&L. Thus the proof is complete.

Proof of Theorem 3.1

We shall only give an outline of the proof because details are similar to those of S&L and L&S. First note the identity

$$\begin{aligned}\hat{x}_t = x_t &- (\hat{\mu}_0 - \mu_0) - (\hat{\mu}_1 - \mu_1)t \\ &- (\hat{\delta}_0 - \delta_0)d_{0t} - (\hat{\delta}_1 - \delta_1)d_{1t}, \quad \text{(A.17)}\end{aligned}$$

which, in conjunction with the assumed consistency properties of the estimators $\tilde{\beta}, \tilde{\alpha}$, and $\tilde{\Omega}$, as well as the consistency results of the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$ obtained from Theorem 2.1, will be central in the subsequent derivations. Using these arguments, one can show that

$$N^{-1} \sum_{t=p+1}^{T} \hat{z}_t \hat{z}_t' = E(z_t z_t') + o_p(1), \qquad \text{(A.18)}$$

$$N^{-3/2} \sum_{t=p+1}^{T} \hat{z}_t \hat{v}_{t-1}' = o_p(1), \qquad \text{(A.19)}$$

and

$$\begin{aligned}&N^{-2} \sum_{t=p+1}^{T} \hat{v}_{t-1} \hat{v}_{t-1}' \\ &= T^{-2} \sum_{t=p+1}^{T} [\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1 - \mu_1)(t-1)] \\ &\quad \times [\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1 - \mu_1)(t-1)]' + o_p(1), \quad \text{(A.20)}\end{aligned}$$

where $z_t = [x_{t-1}'\beta : \Delta x_{t-1}' : \cdots : \Delta x_{t-p+1}']'$ and the first term on the right side of (A.20) converges weakly to $\beta_\perp' C \Omega^{1/2} \int_0^1 \mathbf{B}_*(s) \mathbf{B}_*(s)'\, ds \Omega^{1/2} C' \beta_\perp$ [cf. lemma A (ii) of S&L]. These results can be justified in a straightforward manner by using (A.17) and the consistency properties of the associated estimators. To give a heuristic explanation, note that L&S and S&L obtained (A.18)–(A.20) in a model without any dummy variables. In that case the effect of the estimator $\hat{\mu}_0$ on the left side of (A.18)–(A.20) is asymptotically negligible. Thus, when $a_1 < 1$ in (1.8), the properties of the estimator $\hat{\delta}_1$ are similar to those of $\hat{\mu}_0$ so that it should be clear that the effect of the estimator $\hat{\delta}_1$ on the left side of (A.18)–(A.20) is also asymptotically negligible. For the impulse dummy d_{0t}, this is clear in any case because $\hat{\delta}_0 - \delta_0 = O_p(1)$. When $a_1 = 1$ in (1.8), the step dummy d_{1t} behaves very similarly to the impulse dummy d_{0t}, which explains (A.18)–(A.20) in this case.

Next, note that the error term e_t in (3.1) has the representation

$$\begin{aligned}e_t &= \varepsilon_t - \alpha\beta'(\hat{x}_{t-1} - x_{t-1}) + \Delta\hat{x}_t \\ &\quad - \Delta x_t - \sum_{j=1}^{p-1} \Gamma_j (\Delta\hat{x}_{t-j} - \Delta x_{t-j}) \\ &= \varepsilon_t + \alpha\beta'(\hat{\mu}_0 - \mu_0) \\ &\quad + \alpha\beta'(\hat{\mu}_1 - \mu_1)(t-1) - \Psi(\hat{\mu}_1 - \mu_1) \\ &\quad + \alpha\beta'(\hat{\delta}_0 - \delta_0) d_{0,t-1} \\ &\quad + (\hat{\delta}_0 - \delta_0)\Delta d_{0t} - \sum_{j=1}^{p} \Gamma_j (\hat{\delta}_0 - \delta_0) \Delta d_{0,t-j} \\ &\quad + \alpha\beta'(\hat{\delta}_1 - \delta_1) d_{1,t-1} \\ &\quad + (\hat{\delta}_1 - \delta_1)\Delta d_{1t} - \sum_{j=1}^{p} \Gamma_j (\hat{\delta}_1 - \delta_1) \Delta d_{1,t-j}.\end{aligned}$$

The part of the last expression not involving dummy variables appeared in the work of L&S and S&L, where it was shown that in this case

$$\begin{aligned}&N^{-1} \sum_{t=p+1}^{T} \hat{v}_{t-1} e_t^{*\prime} \tilde{\alpha}_\perp \\ &= N^{-1} \sum_{t=p+1}^{T} [\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1 - \mu_1)(t-1)] \\ &\quad \times [\varepsilon_t' \alpha_\perp C' \beta_\perp - (\hat{\mu}_1 - \mu_1)' \beta_\perp](\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \Psi' \alpha_\perp \\ &\quad + o_p(1). \qquad \text{(A.21)}\end{aligned}$$

Using (A.17) and the consistency properties of the involved estimators, it can similarly be shown that (A.21) holds in the present context. A heuristic explanation can again be obtained by observing that the effect of the estimator $\hat{\mu}_0$ on the left side of (A.21) is asymptotically negligible so that, given the properties of the estimators $\hat{\delta}_0$ and $\hat{\delta}_1$, the same happens also when the impulse dummy d_{0t} and the step dummy d_{1t} are included in the model.

Arguing as in (A.18)–(A.21), it can also be shown that

$$N^{-1/2} \sum_{t=p+1}^{T} \hat{z}_t e_t^{*\prime} \tilde{\alpha}_\perp = O_p(1) \qquad \text{(A.22)}$$

[cf. (A.12) of L&S]. Thus, from (A.18)–(A.22) it follows that we have reduced the problem to that in L&S and S&L so that the stated limiting distribution can be obtained in the same way as in these works.

To see that the RR regression estimators of α, β, and Ω based on (3.1) have the consistency properties stated in Lemma 2.1, we first note that results entirely similar to those in lemmas A3 and A4 of S&L also hold in the present context. In the former case we have to show that the asymptotic behavior of the second sample moments of $\beta'\hat{x}_t, \Delta\hat{x}_t$, and $\beta'_\perp\hat{x}_t$ are similar to those of $\beta'x_t, \Delta x_t$, and $\beta'_\perp x_t$ except that the weak limit of the matrix of second sample moments of $\beta'_\perp\hat{x}_t$ is the same as was obtained in (A.20). Using (A.17) and Theorem 2.1, it is straightforward to show that this is the case and, because the estimation of μ_0 did not affect the previous results, this is also fairly obvious by the heuristic argument used previously. In the case of lemma A4 of S&L, we have to show that the second sample moments between $[\hat{x}'_t\beta : \Delta\hat{x}'_{t-i}]$ and e_t in (3.1) are of order $O_p(T^{-1/2})$ $(i = 1, \ldots, p)$, whereas the second sample moments between $\beta'_\perp\hat{x}_{t-1}$ and e_t are of order $O_p(1)$. Using the representation of e_t given previously and the same arguments as in the preceding case, it can be seen that this result also holds in the present context. After this we can show that the RR regression estimators of α, β, and Ω based on (3.1) have the desired properties. The argument is the same as in lemma A.5 of S&L or lemmas 13.1 and 13.2 of Johansen (1995).

APPENDIX B: DATA SOURCES

Seasonally unadjusted quarterly data for the period from the first quarter of 1975 to the fourth quarter of 1996 (88 observations) were used for the following variables taken from the given sources. The data refer to West Germany until 1990(2) and to the unified Germany afterwards.

Price index: GNP deflator $(1991 = 100)$ from *Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung*. The variable p is the logarithm of the price index.

M3: Nominal monthly values from *Monatsberichte der Deutschen Bundesbank;* the quarterly values are the values of the last month of each quarter. The variable m is log M3 $-p$.

GNP: Real "Bruttosozialprodukt" quarterly values from *Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung*. The variable y is log GNP.

Average bond rate (Umlaufsrendite) (R): Monthly values from *Monatsberichte der Deutschen, Bundesbank;* the quarterly value is the value of the last month of each quarter.

Own rate of M3 (rm): The series was constructed from the interest rates of savings deposits (rs) and the interest rates of three-months time deposits (rt) from *Monatsberichte der Deutschen Bundesbank* as a weighted average as follows:

$$rm = \begin{cases} .24rt + .42rs & \text{for } 1976(1)\text{–}1990(2) \\ .30rt + .33rs & \text{for } 1990(3)\text{–}1996(4) \end{cases}.$$

The weights are chosen according to the relative shares of the corresponding components of M3. The quarterly value is the value of the last month of each quarter.

Interest rate spread: $r = R - rm$.

[Received October 1998. Revised November 1999.]

REFERENCES

Amsler, C., and Lee, J. (1995), "An LM Test for a Unit Root in the Presence of a Structural Change," *Econometric Theory*, 11, 359–368.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10, 271–287.

Campos, J., Ericsson, N. R., and Hendry, D. F. (1996), "Cointegration Tests in the Presence of Structural Breaks," *Journal of Econometrics*, 70, 187–220.

Ghysels, E., and Perron, P. (1996), "The Effects of Linear Filters on Dynamic Time Series With Structural Change," *Journal of Econometrics*, 70, 69–97.

Gregory, A. W., and Hansen, B. E. (1996), "Residual-Based Tests for Cointegration in Models With Regime Shifts," *Journal of Econometrics*, 70, 99–126.

Hackl, P., and Westlund, A. H. (1989), "Statistical Analysis of 'Structural Change': An Annotated Bibliography," *Empirical Economics*, 14, 167–192.

Hansen, B. E. (1992), "Tests for Parameter Instability in Regressions With I(1) Processes," *Journal of Business & Economic Statistics*, 10, 321–335.

Inoue, A. (1999), "Tests of Cointegrating Rank With a Trend-Break," *Journal of Econometrics*, 90, 215–237.

Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, U.K.: Oxford University Press.

Johansen, S., and Nielsen, B. (1993), "Manual for the Simulation Program DisCo," technical report, University of Copenhagen, Institute of Mathematical Statistics.

Johansen, S., and Schaumburg, E. (1999), "Likelihood Analysis of Seasonal Cointegration," *Journal of Econometrics*, 88, 301–339.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Lütkepohl, H., and Saikkonen, P. (2000), "Testing for the Cointegrating Rank of a VAR Process with a Time Trend," *Journal of Econometrics*, 95, 177–198.

Lütkepohl, H., and Wolters, J. (1998), "A Money Demand System for M3 in the Unified Germany," *Empirical Economics*, 23, 371–386.

Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

——— (1990), "Testing for a Unit Root in a Time Series With a Changing Mean," *Journal of Business & Economic Statistics*, 8, 153–162.

Perron, P., and Vogelsang, T. J. (1992), "Nonstationarity and Level Shifts With an Application to Purchasing Power Parity," *Journal of Business & Economic Statistics*, 10, 301–320.

Quintos, C. E. (1998), "Stability Tests in Error Correction Models," *Journal of Econometrics*, 82, 289–315.

Rappoport, P., and Reichlin, L. (1989), "Segmented Trends and Nonstationary Time Series," *Economic Journal*, 99, 168–177.

Saikkonen, P., and Lütkepohl, H. (1999), "Local Power of Likelihood Ratio Tests for the Cointegrating Rank of a VAR Process," *Econometric Theory*, 15, 50–78.

——— (in press), "Trend Adjustment Prior to Testing for the Cointegrating Rank of a VAR Process," *Journal of Time Series Analysis*, 21.

Saikkonen, P., and Luukkonen, R. (1997), "Testing Cointegration in Infinite Order Vector Autoregressive Processes," *Journal of Econometrics*, 81, 93–126.

Seo, B. (1998), "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*, 14, 222–259.

Toda, H. Y. (1994), "Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks When Linear Trends Are Present," *Review of Economics and Statistics*, 76, 66–79.

——— (1995), "Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Ranks in Vector Autoregressions," *Econometric Theory*, 11, 1015–1032.

Wolters, J., Teräsvirta, T., and Lütkepohl, H. (1998), "Modelling the Demand for M3 in the Unified Germany," *Review of Economics and Statistics*, 80, 399–409.

Xiao, Z., and Phillips, P. C. B. (1999), "Efficient Detrending in Cointegrating Regression," *Econometric Theory*, 15, 519–548.

Zivot, E., and Andrews, D. W. K. (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics*, 10, 251–270.

LINKED CITATIONS

Testing for the Cointegrating Rank of a VAR Process with Structural Shifts
Pentti Saikkonen; Helmut Lütkepohl
Journal of Business & Economic Statistics, Vol. 18, No. 4. (Oct., 2000), pp. 451-464.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28200010%2918%3A4%3C451%3ATFTCRO%3E2.0.CO%3B2-2

References

Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence
Anindya Banerjee; Robin L. Lumsdaine; James H. Stock
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 271-287.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C271%3ARASTOT%3E2.0.CO%3B2-C

Tests for Parameter Instability in Regressions with I(1) Processes
Bruce E. Hansen
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 321-335.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C321%3ATFPIIR%3E2.0.CO%3B2-K

The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis
Pierre Perron
Econometrica, Vol. 57, No. 6. (Nov., 1989), pp. 1361-1401.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198911%2957%3A6%3C1361%3ATGCTOP%3E2.0.CO%3B2-W

Testing for a Unit Root in a Time Series with a Changing Mean
Pierre Perron
Journal of Business & Economic Statistics, Vol. 8, No. 2. (Apr., 1990), pp. 153-162.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199004%298%3A2%3C153%3ATFAURI%3E2.0.CO%3B2-0

Nonstationarity and Level Shifts with an Application to Purchasing Power Parity
Pierre Perron; Timothy J. Vogelsang
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 301-320.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C301%3ANALSWA%3E2.0.CO%3B2-L

Segmented Trends and Non-Stationary Time Series
Peter Rappoport; Lucrezia Reichlin
The Economic Journal, Vol. 99, No. 395, Supplement: Conference Papers. (1989), pp. 168-177.
Stable URL:
http://links.jstor.org/sici?sici=0013-0133%281989%2999%3A395%3C168%3ASTANTS%3E2.0.CO%3B2-J

Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks when Linear Trends are Present
Hiro Y. Toda
The Review of Economics and Statistics, Vol. 76, No. 1. (Feb., 1994), pp. 66-79.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199402%2976%3A1%3C66%3AFSPOLR%3E2.0.CO%3B2-O

Modeling the Demand for M3 in the Unified Germany
Jürgen Wolters; Timo Teräsvirta; Helmut Lütkepohl
The Review of Economics and Statistics, Vol. 80, No. 3. (Aug., 1998), pp. 399-409.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199808%2980%3A3%3C399%3AMTDFMI%3E2.0.CO%3B2-R

Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis
Eric Zivot; Donald W. K. Andrews
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 251-270.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C251%3AFEOTGC%3E2.0.CO%3B2-7